REMARKS

The Office Action of December 21, 2001 has been carefully considered. Reconsideration of this application, as amended, is respectfully requested. Claims 1, 7-39 and 41-49 are pending in this application. Of these, claims 1, 39, 43 and 49 are independent. In this Amendment, claims 1, 7-23, 27-29, 34, 39 and 41-42 have been amended, claims 2-6 and 40 have been cancelled, without prejudice, and claims 43-49 have been added.

Changes to the Claims

The amendments to Claims 1 and 39 are amendments submitted to more fully claim that which is applicant's invention, and are not intended to limit or narrow the scope of the claims or to effect the Doctrine of Equivalents as it might be applied to the claims, were they unamended.

Response to Rejections

Claim 1

Pending claim 1 stands rejected under 35 U.S.C. § 102 as being anticipated by Schuetze. The Office Action sets forth the rationale in support of this rejection in Section 3, on pages 2-3.

Applicants respectfully submit that amended independent claim 1 is not anticipated, nor made obvious, by Schuetze. Applicant's claim 1 recites, "the first feature comprising text surrounding an image included in the documents". In contrast to applicants' claim, Schuetze does not teach using a feature related to images included in the documents. Schuetze is silent as to images associated with documents, focusing solely on the co-occurrence of words within documents to determine the similarity between documents. The Office Action asserts that Schuetze discusses images at Column 10, lines 49-53 (See Office Action, page 7, lines 1-6); however, the cited lines read as follows:

> subprocedure, will be described briefly. The first subprocedure, truncated group average agglomerate clustering, merges disjoint document sets, or groups, starting with individuals until only k groups remain. At each step the two groups whose merger would produce the least decrease in average similarity are merged into a single new group.

As Schuetze does not teach or suggest all of the claim limitations, Applicants respectfully requests that the rejection be withdrawn and that claim 1 be allowed.

Claims 7-15

Pending claim 7 stands rejected under 35 U.S.C. § 103 as being made obvious by Scheutze in view of Li. The rationale for this rejection is set forth in the Office Action in Section 4, pages 8-9. Applicants respectfully submit that claim 7 is nonobvious in view of the cited prior art. Each reference discusses the use of a single type of feature, textual features. Neither Schuetze, nor Li, teaches the use of multiple features as claimed by applicants. In contrast, claim 7 represents documents using a first feature corresponding to text surrounding an image and a second feature corresponding to a URL within a document. Further, as discussed above with respect to claim 1, neither reference discloses the use of image related features.

Applicants respectfully submit that claims 8 and 9, which depend from claim 7, are allowable for the reasons discussed with respect to claim 7.

Pending claim 10 also stands rejected under 35 U.S.C. § 103 as being made obvious by Schuetze in view of Li. The rationale for this rejection is set forth in the Office Action in Section 4, pages 8-9. Amended claim 10 recites associating a first and a second feature with the documents being quantitatively represented. Applicants respectfully submit that claim 10 and its dependent claims, claims 11-14, are nonobvious in view of the cited prior for the same reasons discussed above with respect to claims 7-9.

Pending claim 15 also stands rejected under 35 U.S.C. § 103 as being made obvious by Schuetze in view of Li. The rationale for this rejection is set forth in the

Office Action in Section 4, pages 8-9. Amended claim 15 also recites associating a first and a second feature with the documents being quantitatively represented. Applicants respectfully submit that claim 15 is nonobvious in view of the cited prior for the same reasons discussed above with respect to claim 7.

Applicants therefore request that the rejections be withdrawn and claims 7-15 be allowed.

Claims 16-19

Claims 16-17 stand rejected under 35 U.S.C. § 102 as being anticipated by Schuetze. The Office Action sets forth the rationale for this rejection in Section 3, pages 2-6. Applicants respectfully submit that claims 16 and 17 are not anticipated, nor made obvious, by Schuetze because they depend from claims 15 and 1, both of which patentably define over the cited prior art.

Claims 18-19 stand rejected under 35 U.S.C. § 103 as being made obvious by Schuetze in view of Li. The rationale for this rejection is set forth in the Office Action in Section 4, pages 8-9. Applicants respectfully submit that claims 18 and 19 are not made obvious by Schuetze in view of Li because they depend from claims 15 and 1, both of which patentably define over the cited prior art.

Applicants therefore request that the rejections be withdrawn and claims 16-19 be allowed.

Claims 20-21 and New Independent Claim 49

Claim 20 stands rejected under 35 U.S.C. § 102 as being anticipated by Schuetze. The Office Action sets forth the rationale for this rejection in Section 3, pages 2-6. Applicants respectfully submit that amended claim 20 patentably defines over the cited prior art. Claim 20 now depends from new, independent claim 49, which recites a method of representing documents in feature space using a first feature and a second feature. The first feature is an image feature and the second feature may be any one of a set of multi-modal features including a text feature, a hyperlink feature and a genre feature. Independent claim 49 and its dependent claim 20 patentably define over Schuetze and Li for at least three reasons. First, neither reference discloses, nor makes obvious, the use of multiple features of differing modes, as recited by applicants' claim 49. Second, neither reference discloses, nor suggests, the use of a genre feature to quantitatively represent documents. As discussed in the Application at page 21, line 23 – page 22, line 8:

> A document genre is a culturally defined document category that guides a document's interpretation. Genres are signaled by the greater document environment (such as the physical media, pictures, titles, etc. that serve to distinguish at a glance, for example, the National Enquirer from the New York Times) rather than the document text. The same information presented in two different genres may lead to two different interpretations. For example, a document starting with the line "At dawn the street was peaceful . . ." would be interpreted differently by a reader of Time Magazine than by a reader of a novel. Each document type has an easily recognized and culturally defined genre structure which guides our understanding and interpretation of the information it contains. For example, news reports, newspaper editorials, calendars, press releases, and short stories are all examples of possible genres. A document's structure and genre can frequently be determined (at least in part) by an automated analysis of the document or text (step 510).

Finally, Applicants submit that claim 49 and dependent claim 20 define over the cited prior art because of the use of an image feature, as previously discussed above with respect to claim 1.

Applicants respectfully submit that amended pending claim 21 is allowable over the cited prior art because it depends from claims 20 and 49, which patentably define over the cited prior art as discussed above.

Applicants therefore request that the rejections be withdrawn and claims 20-21 and 49 be allowed.

Claims 22-38

Pending claim 22 stands rejected under 35 U.S.C. § 102 as being anticipated by Schuetze. The Office Action sets forth the rationale for this rejection in Section 3, page 7, lines 1-6, there purporting to find support in column 10, lines 49-53 of Schuetze. Applicants respectfully submit that Schuetze at column 10, lines 49-53, cannot reasonably be interpreted as disclosing the use of image based features, let alone the use of a color histogram for an image. At the cited location Schuetze states:

> subprocedure, will be described briefly. The first subprocedure, truncated group average agglomerate clustering, merges disjoint document sets, or groups, starting with individuals until only k groups remain. At each step the two groups whose merger would produce the least decrease in average similarity are merged into a single new group.

Applicants further submit that claim 22 is allowable over the cited prior art to the reasons discussed above with respect to claim 49, from which claim 22 depends.

Applicants also submit that pending claims 23-27 are allowable over the cited prior art because they depend from claim 22, which defines patentable subject matter.

Claim 28 stands rejected under 35 U.S.C. § 102 as being anticipated by Schuetze. The Office Action sets forth the rationale for this rejection in Section 3, page 7, lines 7-8, again claiming to find support at column 10, lines 49-53 of Schuetze. Applicants again respectfully submit that Schuetze at column 10, line 49-53, cannot fairly be characterized as using image related features to quantitatively represent documents, as discussed previously with respect to independent claims 1 and 49. Applicants therefore submit that claim 28 is allowable over the cited prior art as are claims 29-38, which depend from claim 28.

Applicants therefore request that the rejections be withdrawn and claims 22-38 be allowed.

Claims 39-42

Claims 39-40 stand rejected under 35 U.S.C. § 102 as being anticipated by Schuetze. The Office Action sets forth the rationale for this rejection in Section 3, page 7, lines 9-17. Claim 39 has been rewritten as an independent claim and includes limitations of claim 1, from which it previously depended, and claim 40. Applicants respectfully submit that amended claim 39 is not anticipated, nor made obvious by Schuetze because Schuetze does not teach using users' document usage histories to represent the documents themselves. While the word "user" does appear in Schuetze at column 13, line 66, in the following discussion in column 14 Schuetze deals with strictly textual features to a organize documents into clusters, rather than users' document usage histories.

Applicants respectfully submit that claims 41-42, which depend from claim 39 are also allowable over the cited prior art.

Applicants therefore request that the rejections be withdrawn and claims 39-42 be allowed.

Claims 43-48

New claims 43-48 are presented for examination in this paper. Independent claim 43 is directed to a computer-readable medium storing instructions for quantitatively representing documents using a first feature and a second feature. The second feature is one of a set of multi-modal features. As discussed previously, the use of multiple, multi-modal features to represent documents is believed to patentably define over the cited prior art.

New claim 44 depends from claim 43 and recites that the first feature is an image feature.

New claim 45 depends from claim 44 and recites that the first feature comprises a color histogram for an image included in a document. As such, claim 45 is similar claim 22 prior to its amendment.

New claim 46 depends from claim 44 and is similar in subject matter to claim 23.

New claim 47 depends from claim 44 and includes subject matter similar to that of claim 28.

New claim 48 depends from claim 47 and includes subject matter similar to that of claim 29.

Applicant therefore requests that the claims be allowed.

Reconsideration/Admittance Requested

In view of the foregoing remarks and amendments, reconsideration of this application and allowance thereof are earnestly solicited.

Fee Authorization And Extension Of Time Statement

A two month extension of time believed to be required for this amendment. The undersigned Xerox Corporation attorney hereby authorizes the charging of any necessary fees, other than the issue fee, to Xerox Corporation Deposit Account No. 24-0025. This also constitutes a request for the needed extension of time and authorization to charge all fees therefor to Xerox Corporation Deposit Account No. 24-0025.

In the event the Examiner considers personal contact advantageous to the disposition of this case, he is hereby authorized to call Applicant's attorney, Nola Mae McBain, at Telephone Number (650) 812-4264, Palo Alto, California.

Respectfully submitted,

Nola Mae McBain

Nola Mae McBain
Attorney for Applicant(s)
Registration No. 35,782
Telephone: 650-812-4264

Date: May 17, 2002

APPENDIX A

Marked Up Amended Claims Under 37 C.F.R. 1.121(b)(1)(iii):

Appendix A sets forth a marked up version of the prior pending amended claims for their corresponding pending claims with additions shown with underlining (e.g. new text) and deletions shown with a strikethrough (e.g. ~~delete text~~).

1. (Amended). A method for quantitatively representing ~~objects~~ documents in a vector space, comprising the steps of:

identifying ~~an object~~ a first document to be processed from a plurality of ~~objects~~ documents;

extracting a first feature corresponding to the ~~object~~ first document from the plurality of ~~objects~~ documents, the first feature comprising text surrounding an image included in the document;

converting the first feature to ~~at least one~~ a first vector; and

associating the ~~at least one~~ first vector with the ~~object~~ first document.

Claims 2-6 have been cancelled.

7. (Amended) The method of claim 1 ~~2, wherein the feature comprises the subject document in the collection of documents.~~ further comprising the steps of:

extracting a second feature corresponding to the document, the second feature comprising a first URL representing the first document;

converting the second feature to a second vector; and

associating the second vector with the first document .

8. (Amended) The method of claim 7, wherein the ~~converting~~ step of converting the second feature comprises the sub-steps of:

identifying each unique word within the URLs representing all documents in the collection of documents; and

counting the occurrences of each unique word in the ~~subject document~~ first URL;

creating a vector having a number of dimensions equal to the number of unique words in the URLs representing all documents in the collection of documents, and further having as each element a numeric value representative of the number of occurrences in the ~~subject document~~ first URL of the corresponding word.

9. (Amended) The method of claim 8, wherein the value representative of the number of occurrences in the ~~subject document~~ first URL of the corresponding word is calculated as the token frequency weight of the corresponding word multiplied by the inverse context frequency weight of the corresponding word.

10. (Amended) The method of claim 1 ~~2, wherein the feature comprises inlinks in the collection of documents linking to the subject document.~~ further comprising the steps of:

extracting a second feature corresponding to the first document, the second feature comprising inlinks in the collection of documents linking to the first document;

converting the second feature to a second vector; and

associating the second vector with the first document .

11. (Amended) The method of claim 10, wherein the ~~converting~~ step of converting the second feature comprises the sub-steps of:

identifying each document having links within the collection of documents;

determining how many times each document having links points to the ~~subject~~ first document; and

creating ~~a~~ the second vector ~~having~~ a number of dimensions equal to the number of documents having links in the collection of documents, and the second vector further having as each element a numeric value representative of the number of links in each corresponding document linking to the ~~subject~~ first document.

12. (Amended) The method of claim 11, wherein the numeric value representative of the number of links in each corresponding document linking to the ~~subject~~ first document is calculated as the token frequency weight of the corresponding link multiplied by the inverse context frequency weight of the corresponding link.

13. (Amended) The method of claim 10, wherein the ~~converting~~ step of converting the second feature comprises the sub-steps of:

identifying each document having hyperlinks within the collection of documents, and further identifying each unique word associated with URLs defining hyperlinks in each document;

counting the occurrences of each unique word in the URLs defining hyperlinks pointing to the ~~subject~~ first document; and

creating ~~a~~ the second vector having a number of dimensions equal to the number of unique words associated with URLs defining hyperlinks within the collection of documents, and the second vector further having as each element a numeric value representative of the number of occurrences in the URLs defining hyperlinks pointing to the ~~subject~~ first document of the corresponding word.

14. (Amended) The method of claim 13, wherein the numeric value representative of the number of occurrences in the URLs defining hyperlinks pointing to the ~~subject~~ first document of the corresponding word is calculated as the token frequency weight of the corresponding word multiplied by the inverse context frequency weight of the corresponding word.

15. The method of claim 1 ~~2, wherein the feature comprises outlinks in the subject document linking to other documents.~~ further comprising the steps of:

extracting a second feature corresponding to the first document, the second feature comprising outlinks in the collection of documents linking to the first document;

converting the second feature to a second vector; and associating the second vector with the first document .

16. (Amended) The method of claim 15, wherein the ~~converting~~ step of converting the second feature comprises the sub-steps of:

identifying each other document linked to by all documents within the collection of documents; and

creating ~~a~~ the second vector having a number of dimensions equal to the number of other documents linked to by documents in the collection of documents, and the second vector further having as each element a numeric value representative of the number of links in the ~~subject~~ first document linking to each corresponding other document.

17. (Amended) The method of claim 16, wherein the numeric value representative of the number of links in the ~~subject~~ first document linking to each corresponding other document is calculated as the token frequency weight of the corresponding link multiplied by the inverse context frequency weight of the corresponding link.

18. (Amended) The method of claim 15, wherein the ~~converting~~ step of converting the second feature comprises the sub-steps of:

identifying each unique word associated with URLs defining hyperlinks in each document in the collection of documents;

counting the occurrences of each unique word in the URLs defining hyperlinks in the ~~subject~~ first document; and

creating ~~a~~ the second vector having a number of dimensions equal to the number of unique words associated with the URLs defining hyperlinks in each document, and the second vector further having as each element a numeric value representative of the number of occurrences in the URLs defining hyperlinks in the ~~subject~~ first document of the corresponding word.

19. (Amended) The method of claim 18, wherein the numeric value representative of the number of occurrences in the URLs defining hyperlinks in the ~~subject~~ first document of the corresponding word is calculated as the token frequency weight of the corresponding word multiplied by the inverse context frequency weight of the corresponding word.

20. (Amended) The method of claim ~~2~~ 49, wherein the second feature comprises ~~the~~ a text genre ~~of the text represented by the subject document~~ feature.

21. (Amended) The method of claim 20, wherein the ~~converting~~ step of converting the second feature comprises the sub-steps of:

for each possible text genre, processing the ~~subject~~ first document to calculate the probability that the ~~subject~~ first document is of the corresponding text genre; and

creating ~~a~~ the second vector having a number of dimensions equal to the number of possible text genres, and the second vector further having as each element a numeric value representative of the probability that the ~~subject~~ first document is of the corresponding genre.

22. (Amended) The method of claim ~~2~~ 49, wherein the first feature comprises the color histogram for an image represented by the ~~subject~~ first document.

23. (Amended) The method of claim 22, wherein the ~~converting~~ step of converting the first feature comprises the sub-steps of:

quantizing the image represented by the ~~subject~~ first document into a multi-dimensional color model;

creating a color histogram having a plurality of bins for each dimension in the color model, each bin corresponding to a unique combination of binary bits representing information from the associated dimension of the color model;

counting each of a plurality of pixels from the image in a corresponding bin associated with each dimension of the color model; and creating ~~a~~ the first vector having a number of dimensions equal to the total number of bins in the color histogram, and the first vector further having as each element a numeric value representative of the number of pixels in the image corresponding to the corresponding histogram bin.

27. (Amended) The method of claim 23, wherein the image represented by the ~~subject~~ first document comprises a region of a bitmap.

28. (Amended) The method of claim ~~2~~ 49, wherein the first feature comprises the color complexity of an image represented by the ~~subject~~ first document.

29. (Amended) The method of claim 28, wherein the ~~converting~~ step of converting the first feature comprises the sub-steps of:

quantizing the image represented by the ~~subject~~ first document into a multi-dimensional color model;

determining the maximum number of pixels in any row in any image represented by ~~a~~ any document in the collection of documents;

determining the maximum number of pixels in any column in any image represented by ~~a~~ any document in the collection of documents;

creating a horizontal complexity histogram and a vertical complexity histogram, each having a number of bins equal to the maximum number of pixels in any row and in any column, respectively;

identifying horizontal runs of pixels of all possible lengths in the quantized image, and for each possible length, counting the number of pixels in a plurality of rows of the quantized image belonging to the horizontal runs in a corresponding bin of the horizontal complexity histogram;

identifying vertical runs of pixels of all possible lengths in the quantized image, and for each possible length, counting the number of pixels in a plurality of columns of the quantized image belonging to the vertical runs in a corresponding bin of the horizontal complexity histogram;

creating a horizontal complexity vector having a number of dimensions equal to the maximum number of pixels in any row, and further having as each element a numeric value representing the number of pixels in the image in the corresponding horizontal histogram bin; and

creating a vertical complexity vector having a number of dimensions equal to the maximum number of pixels in any column, and further having as each element a numeric value representing the number of pixels in the image in the corresponding vertical histogram bin.

34. (Amended) The method of claim 28, wherein the ~~converting~~ step of converting the first feature comprises the sub-steps of:

quantizing the image represented by the ~~subject~~ first document into a multi-dimensional color model;

determining the maximum number of pixels in any row in any image represented by ~~a~~ any document in the collection of documents;

determining the maximum number of pixels in any column in any image represented by ~~a~~ any document in the collection of documents;

creating a horizontal complexity histogram and a vertical complexity histogram, each having a selected number of bins corresponding to a plurality of quantized ranges of run lengths;

identifying horizontal runs of pixels of all possible lengths in the quantized image, and for each possible length, counting the number of pixels in a plurality of rows of the quantized image belonging to the horizontal runs in a corresponding bin of the horizontal complexity histogram;

identifying vertical runs of pixels of all possible lengths in the quantized image, and for each possible length, counting the number of pixels in a plurality of columns of the quantized image belonging to the vertical runs in a corresponding bin of the horizontal complexity histogram;

creating a horizontal complexity vector having a number of dimensions equal to the selected number of bins in the horizontal complexity histogram, and further having as each element a numeric value representing the number of pixels in the image in the corresponding horizontal histogram bin; and

creating a vertical complexity vector having a number of dimensions equal to the number of bins in the vertical complexity histogram, and further having as each element a numeric value representing the number of pixels in the image in the corresponding vertical histogram bin.

39. (Amended) ~~The method of claim 1, wherein the object to be processed comprises a subject user selected from a user population.~~ A method for quantitatively representing in a vector space users of a collection of documents, comprising the steps of:

identifying a first user to be processed from the users of the collection of documents;

extracting from the collection of documents a first feature representing a first sub-set of documents of the collection that have been accessed by the first user;

converting the first feature to a first vector; and

associating the first vector with the first user.

40. Claim 40 has been cancelled.

41. (Amended) The method of claim ~~40~~ 39, wherein the converting step comprises the steps of:

identifying each unique document in the collection of documents;

calculating the number of times the ~~subject~~ first user accessed each document in the collection of documents; and

creating ~~a~~ the first vector having a number of dimensions equal to the number of documents in the collection of documents, and the first vector further having as each element a numeric value representative of the number of times the ~~subject~~ first user has accessed the corresponding document.

42. (Amended) The method of claim 41, wherein the value representative of the number of times the ~~subject~~ first user has accessed the corresponding document is calculated as the token frequency weight of the corresponding document multiplied by the inverse context frequency weight of the corresponding document.

Claims 43 – 49 have been added.